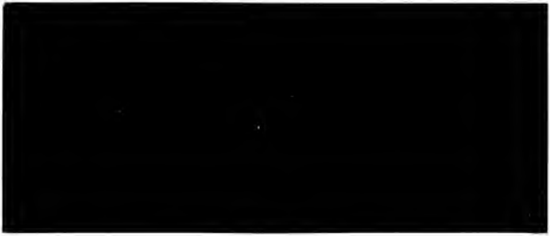
LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

THE VARIANCES OF REGRESSION COEFFICIENT ESTIMATES USING AGGREGATE DATA*

by

Roy E. Welsch and Edwin Kuh

Working Paper 616-72
October 1972



DEC 7 197
DEWEY LIBRARY

# THE VARIANCES OF REGRESSION COEFFICIENT ESTIMATES USING AGGREGATE DATA*

by

Roy E. Welsch and Edwin Kuh

Working Paper 616-72
October 1972

*This research was supported in part by National Science Foundation Grant GJ1154X.

RECEIVED
JAN 9 1973
LIBRARIES

ABSTRACT

Regression analysis based on aggregative data is sometimes thought to be inferior to the same model fitted to micro data. The latter, however, is often unavailable. In this paper we derive bounds for the variances of the macro parameter estimates for a linear regression model with random coefficients and an additive error term and show that, under plausible conditions, the macro parameter variances will diminish as the members comprising the aggregate increase. It is also shown that the Theil aggregation weights and more readily obtainable proportions (to which the weights are usually related) must decrease in size as the aggregate population is increased for the same result to hold. Under circumstances that often exist in practice, the macro estimates will be quite efficient ones.


C3C653

## 1. Introduction

Let $\underline{\beta}_{it}$ denote the vector at time t $(t=1,2,\ldots,T)$ of unobservable parameters of the micro regression structures $y_{it} = \underline{x}_{it}\underline{\beta}_{it} + \varepsilon_{it}$, $i=1,2,\ldots,N$, associated with N individuals. We shall assume that these micro regression parameters vary with time according to a wide-sense stationary stochastic process with $E(\underline{\beta}_{it}) = \underline{\beta}$. This paper considers the estimation of the vector $\underline{\beta}$ using the macro data $Y_t = \sum_{i=1}^{N} y_{it}$ and $\underline{X}_t = \sum_{i=1}^{N} \underline{x}_{it}$.

This type of problem occurs frequently in economics and other social sciences where some or all of the micro data $(\underline{x}_{it}, y_{it})$ has not been recorded and only the macro (aggregate) data is available. The absence of micro data is not the only case when we might want to consider aggregation. Given the micro data, one way to estimate $\underline{\beta}$ would be to average the estimates obtained from each micro equation. For reasonable N this involves a lot of computation, and perhaps an estimate based on the aggregate data would be about as good.

In an earlier paper (Kuh (1972)) it was shown under certain reasonable assumptions, that the variances of the estimated macro coefficients decrease as the number of individuals in the aggregate increase. Empirical tests of the assumptions supported the theoretical propositions. The random coefficient assumption used there differs from the one on this paper, in that randomness was conceived to be cross-sectional in nature: an aggregate relation may include heterogeneous micro parameters which can be thought of as random drawings from a population with a stable mean. Here, randomness is treated as a time-series, stationary stochastic process where each micro-coefficient vector is assumed to have the same mean. From a formal point

of view the sources of randomness are very different, but in both cases, aggregation gain, defined as reduction in the estimated macro variance relative to the micro variances can, and often will occur. Suitable aggregation, i.e. aggregation for which the mean population parameters (as well as the random process laws) behave approximately as assumed, thus has a valid justification which heretofore has been absent. Casual agglomeration of data, however, where in particular the population micro parameters are not stable as assumed, will evidently cause trouble. A subsequent paper using empirical data and Monte Carlo experiments will explore more fully the implications of these results for empirical research, some of which have already been drawn in Kuh (1972).

The results presented are similar to those of Theil (1968), who only treated variability in the pure random macro population parameter case, whereas we treat a more complete model as well as estimation properties. Furthermore, at the end of this paper, the relation between the aggregation weights of Theil (1954) and aggregation is shown. Empirical inferences about aggregation weights is available (see Kuh (1972)) from individual shares in the aggregate that cast some light on the question of aggregation gain.

The next three sections of the paper present the basic regression model and derive bounds for the macrovariances. In section 5, the limiting properties of the macrovariance using the bounds derived earlier show what conditions must hold for aggregation gain to occur.

## 2. The Regression Model

For the micro equations, we assume the existence of the regression structures

(2.1) $$y_{it} = \underline{x}_{it}\underline{\beta}_{it} + \varepsilon_{it} \qquad \begin{matrix} i=1,2,\ldots,N \\ t=1,2,\ldots,T \end{matrix}$$

where

$y_{it}$ is the dependent variable,

$\underline{x}_{it}$ is a 1 x K vector of "explanatory" variables,

$\underline{\beta}_{it}$ is a K x 1 vector of unknown regression parameters,

$\varepsilon_{it}$ is the additive "error" component,

and

a. $E(\varepsilon_{it}) = 0$

b. $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_i^2\delta_{st}\ \delta_{ij}$ where $\delta$ is the Kronecker delta and the $\sigma_i^2$ are scalars with $\sup_i \sigma_i^2 < \sigma^2$.

(2.2) c. $\underline{\beta}_{it}$ is the realization of a multivariate wide-sense stationary stochastic process with $E(\underline{\beta}_{it}) = \underline{\beta}$ and $E(\underline{\beta}_{it}-\underline{\beta})(\underline{\beta}_{it+s}-\underline{\beta})' = r_i(s)$

d. The processes $\underline{\beta}_{it}$ and $\underline{\beta}_{jt}$ are uncorrelated for $i \neq j$

e. $\underline{\beta}_{it}$ and $\varepsilon_{js}$ are uncorrelated for all s and t and i and j.

These conditions imply that we have N uncorrelated regression equations where the parameters in each equation are realizations of a stochastic process with the same mean vector. For each i this model is identifical to that proposed by Burnett and Guthrie (1970).

The above assumptions require comment. (2.2a) is a standard assumption in regression analysis that we retain here. (2.2b) permits the additive error variance to differ among individuals, but these errors are postulated to be independent across micro units. To the extent that they are not, efficiency gains from generalized least squares estimation which demand all micro data be used, will be foregone [Swamy (1970)]. (2.2c) allows for what, in principle, could be a complex autocorrelated random process in the micro coefficients. Relaxation of the assumption that the population micro parameters are fixed for all time permits a substantial increase in realism. Individual firms or persons often behave according to a stable underlying process, but that behavior often departs from its basic (i.e. average) modus operandi in more complicated ways than are represented by an additive error term. (2.2d) asserts that random coefficient processes across individuals are uncorrelated. We have somewhat less faith in the validity of this assumption than the others. Oligopoly interdependence in particular could lead to violations of this assumption. In general, however, there does not appear to be greater departures from reality in this instance than in others made in the estimation of economic or social behavior relationships. Finally, (2.1e) asserts that the two sources of randomness are uncorrelated, a proposition that is convenient and does not appear to be a particular cause for concern.

In summary, the random coefficient model allows for much richer behavioral variations that should be considered in an aggregation context. Since there are two sources of random variation assumed to be independent, results from the following analysis hold for either alone, or both, so the reader can choose which aspects are most appealing for his immediate estimation concerns.

Let

$$(2.3) \qquad \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} \qquad\qquad \underline{\underline{X}} = \begin{bmatrix} \underline{X}_1 \\ \vdots \\ \underline{X}_T \end{bmatrix}$$

and assume that the T x K matrix $\underline{\underline{X}}$ is of full rank. We propose to estimate $\underline{\beta}$ by

$$(2.4) \qquad \underline{b} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'\underline{Y} .$$

This estimator is unbiased because

$$(2.5) \qquad E(Y_t) = \sum_{i=1}^{N} E(y_{it}) = \sum_{i=1}^{N} \underline{x}_{it} E(\underline{\beta}_{it}) + \sum_{i=1}^{N} E(\varepsilon_{it})$$

$$= \sum_{i=1}^{N} \underline{x}_{it}\underline{\beta} = \underline{X}_t\underline{\beta}$$

which implies that $E(\underline{Y}) = \underline{\underline{X}}\underline{\beta}$.

## 3. Variance Properties of the Aggregate Estimator

For each N we can compute the covariance matrix of $\underline{b}$, denoted by $V_N(\underline{b})$. We are interested in finding conditions so that the elements of $V_N(\underline{b})$ will remain bounded (or go to zero) as N increases. Let $\underline{\underline{z}}_i$ be the T x KT matrix

$$\underline{\underline{z}}_i = \begin{bmatrix} \underline{x}_{i1} & \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{x}_{i2} & \underline{0} & \cdots & \vdots \\ \underline{0} & \underline{0} & \underline{x}_{i3} & \underline{0} & \vdots \\ \cdots & \cdots & \underline{0} & \ddots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \underline{x}_{iT} \end{bmatrix}$$

and set

$$\underline{\beta}_i = \begin{bmatrix} \underline{\beta}_{i1} \\ \vdots \\ \underline{\beta}_{iT} \end{bmatrix} \qquad \underline{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

$$\underline{\underline{G}} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'.$$

Then

$$(3.1) \qquad \underline{Y} = \sum_{i=1}^{N} (\underline{\underline{z}}_i\underline{\beta}_i + \underline{\varepsilon}_i)$$

and

$$(3.2) \qquad (\underline{b}-\underline{\beta}) = \underline{\underline{G}} \left[ \sum_{i=1}^{N} \underline{\underline{z}}_i(\underline{\beta}_i-\underline{\beta}^*) + \underline{\varepsilon}_i \right]$$

where $\underline{\beta}^*$ is the KT x 1 vector

$$\begin{bmatrix} \underline{\beta} \\ \vdots \\ \underline{\beta} \end{bmatrix} .$$

It is now convenient to define the KT x KT matrix

$$\underline{\underline{V}}_i = \begin{bmatrix} \underline{\underline{r}}_i(0) & \underline{\underline{r}}_i(1) & \underline{\underline{r}}_i(T-1) \\ \vdots & \vdots & \vdots \\ \underline{\underline{r}}_i(T-1) & \underline{\underline{r}}_i(T-2) & \underline{\underline{r}}_i(0) \end{bmatrix}$$

which represents the covariance structure of the stochastic process described in (2.2c). If we use assumptions b, c, d, e of (2.2) then

$$(3.3) \qquad \underline{\underline{V}}_N(\underline{b}) = E(\underline{b}-\underline{\beta})(\underline{b}-\underline{\beta})' = \sum_{i=1}^{N} \underline{\underline{Gz}}_i \, \underline{\underline{V}}_i \underline{\underline{z}}_i' \, \underline{\underline{G}}' + \sum_{i=1}^{N} \sigma_i^2 (\underline{\underline{X}}'\underline{\underline{X}})^{-1} .$$

Since $V_N(\underline{b})$ is a covariance matrix, the Cauchy-Schwarz inequality implies that in order to make estimates about the magnitude of the elements of $\underline{\underline{V}}_N(b)$ we need only examine the diagonal elements.

<u>Theorem</u> 1. If

(3.4)

(a) $\sup\limits_{\substack{\text{all } i \\ 1 \le \ell \le K \\ 1 \le t \le T}} |x_{i\ell t}| \le M$

(b) $\sup\limits_{\substack{\text{all } i \\ 1 \le p \le TK \\ 1 \le q \le TK}} |V_{ipq}| \le M'$

then

$$(3.5) \qquad [V_N(b)]_{\ell\ell} \leq (\lambda T+\sigma^2)N(\underline{X}'\underline{X})^{-1}_{\ell\ell}$$

where $\lambda$ is a constant independent of N.

Proof. Let $|g^{(\ell)}|$ denote the vector whose components are the absolute value of the components of $g^{(\ell)}$, the $\ell^{\text{th}}$ row of $(\underline{X}'\underline{X})^{-1}\underline{X}'$, and $\underline{1}$ denote the T x T matrix with each component equal to 1. Conditions (a) and (b) of (3.4) imply that

$$(3.6) \qquad \sup_{\substack{\text{all } i\\ 1\leq p\leq T\\ 1\leq q\leq T}} |(\underline{Z}_i\underline{V}_i\underline{Z}_i')_{pq}| \leq \lambda < \infty$$

Now

$$(3.7) \qquad (\underline{V}_N(\underline{b}))_{\ell\ell} = \sum_{i=1}^{N} \underline{g}^{(\ell)}\underline{Z}_i\underline{V}_i\underline{Z}_i'\underline{g}'^{(\ell)} + \sum_{i=1}^{N} \sigma_i^2 (\underline{X}'\underline{X})^{-1}_{\ell\ell}$$

and

$$(3.8) \qquad \sum_{i=1}^{N} \underline{g}^{(\ell)}\underline{Z}_i\underline{V}_i\underline{Z}_i'\underline{g}'^{(\ell)} = \text{tr} \sum_{i=1}^{N} \underline{Z}_i\underline{V}_i\underline{Z}_i'\underline{g}'^{(\ell)}\underline{g}^{(\ell)}$$

$$\leq N\lambda \text{ tr } \underline{1}|\underline{g}'^{(\ell)}||\underline{g}^{(\ell)}|$$

But by the Cauchy-Schwarz inequality

$$(3.9) \qquad \text{tr } \underline{1}|\underline{g}'^{(\ell)}||\underline{g}^{(\ell)}| = [\sum_{j=1}^{T} |g_j^{(\ell)}|]^2 \leq T \sum_{j=1}^{T} (g_j^{(\ell)})^2$$

and

$$(3.10) \qquad \sum_{j=1}^{T} (g_j^{(\ell)})^2 = |\underline{g}^{(\ell)}||g'^{(\ell)}| = (\underline{X}'\underline{X})^{-1}_{\ell\ell} .$$

The inequality (3.5) follows since (2.2b) implies that

$$\sum_{i=1}^{N} \sigma_i^2 \leq N\sigma^2.$$

Thus we have shown that under rather plausible conditions we can examine $V_N(\underline{b})$ by looking at $N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$. Condition (3.4a) merely states that all elements of the explanatory variables should be bounded and (3.4b) imposes a mild restriction on the covariance structure of the stochastic processes generating the $\underline{\beta}_{it}$.

## 4. The Structure of $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$.

From the above discussion it is possible to see that $N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ plays the crucial role in determining the gains that might be obtained from aggregation. We can always compute $N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ but it is useful to see what conditions imposed on $N(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ imply about the structure of the micro regression equations.

Let $\underline{X}^{(\ell)}$ $\ell=0,1,\ldots,K-1$ denote the columns of $\underline{\underline{X}}$. It is convenient to assume that the first element of $\underline{x}_{it}$ is 1 for all i and t. Therefore $\underline{X}^{(0)}$ is a column vector with each element equal to N. Finally let $\sigma_\ell^2$ be the variance of $\underline{X}^{(\ell)}$ (we assume $\sigma_\ell^2 > 0$, $\ell=1,\ldots,K-1$), and $\sigma_{\ell\cdot\cdot}^2$ the error variance for the regression of $\underline{X}^{(\ell)}$ on the remaining K-2 explanatory macrovariables.

Theorem 2. If (3.4a) holds, then

$$(4.1)\qquad (\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{\ell\ell} = \frac{1}{T\sigma_{\ell\cdot\cdot}^2} \qquad \ell=1,2,\ldots,K-1$$

and

$$(4.2)\qquad (\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{00} \le \frac{1}{N^2T} + \frac{M^2K^2}{T \min\limits_{1\le\ell\le K-1} \sigma_{\ell\cdot\cdot}^2} .$$

Proof. In what follows $\rho_{\ell m}$ stands for the simple correlation between macro explanatory variables $\underline{X}_\ell$ and $\underline{X}_m$ and $V_\ell$ denotes the coefficient of variation of the macro variable $\underline{X}_\ell$.

First we note that

$$(4.3)\qquad [\underline{\underline{X}}'\underline{\underline{X}}] = \begin{bmatrix} N^2T & \underline{X}_0'\cdot\underline{X}_1 & \cdots & \underline{X}_0'\cdot\underline{X}_{K-1} \\ \underline{X}_0'\cdot\underline{X}_1 & \underline{X}_1'\cdot\underline{X}_1 & \ddots & \\ \vdots & & \ddots & \\ \underline{X}_0'\cdot\underline{X}_{K-1} & \cdots & \cdots & \underline{X}_{K-1}'\cdot\underline{X}_{K-1} \end{bmatrix} .$$

By elementary row and column operations $[\underline{\underline{X}}'\underline{\underline{X}}]$ can be transformed into a partitioned matrix whose lower right hand partition is the correlation matrix $[\underline{\underline{C}}]$.

$$(4.4)\qquad [\underline{\underline{X}}'\underline{\underline{X}}] = [\underline{\underline{F}}][\underline{\underline{G}}]\left[\begin{array}{c|ccccc} N & 1/V_1 & 1/V_2 & \cdots & & 1/V_{K-1} \\ \hline 0 & 1 & \rho_{12} & \cdots & & \rho_{1,K-1} \\ \vdots & & & & & \\ \vdots & \rho_{12} & 1 & \ddots & & \vdots \\ \vdots & & & \ddots & & \vdots \\ 0 & \rho_{1,K-1} & \cdots & & \cdot & 1 \end{array}\right][\underline{\underline{G}}]$$

$$= [\underline{\underline{F}}][\underline{\underline{G}}]\left[\begin{array}{c|ccc} N & 1/V_1 & \cdots & 1/V_{K-1} \\ \hline 0 & & & \\ \vdots & & \underline{\underline{C}} & \\ \vdots & & & \end{array}\right][\underline{\underline{G}}]$$

where

$$(4.5)\qquad [\underline{\underline{F}}] = \begin{bmatrix} NT & & & & \underline{\underline{0}} \\ \underline{1}\cdot\underline{X}_1 & T & & & \\ \underline{1}\cdot\underline{X}_2 & 0 & T & & \\ & & & \ddots & \\ \underline{1}\cdot\underline{X}_{K-1} & 0 & \cdots & \cdots & T \end{bmatrix}$$

is a product of elementary row operations that reverse the subtraction of means from second moments about the origin and

$$(4.6) \qquad [\underline{\underline{G}}] = \begin{bmatrix} 1 & & & & \\ & \sigma_1 & & \underline{\underline{0}} & \\ & & \sigma_2 & & \\ & & & \ddots & \\ & \underline{\underline{0}}' & & & \sigma_{K-1} \end{bmatrix}$$

is used for elementary row and column operations to reverse the conversion of centered moments into correlations. If $\underline{\underline{D}} = \underline{\underline{C}}^{-1}$, then

$$(4.7) \qquad [\underline{X}'\underline{X}]^{-1} = [\underline{\underline{G}}]^{-1} \left[\begin{array}{c|cccc} 1/N & \alpha_1 & \alpha_2 & \cdots & \alpha_{K-1} \\ \hline 0 & & & & \\ \vdots & & & \underline{\underline{D}} & \\ 0 & & & & \end{array}\right] [\underline{\underline{G}}]^{-1}[\underline{\underline{F}}]^{-1}$$

where

$$(4.8) \qquad [\underline{\underline{F}}]^{-1} = \begin{bmatrix} 1/NT & & & & \\ -\frac{\underline{1}\cdot\underline{X}_1}{NT^2} & \frac{1}{T} & & \underline{\underline{0}} & \\ -\frac{\underline{1}\cdot\underline{X}_2}{NT^2} & 0 & \frac{1}{T} & & \\ \vdots & & & \ddots & \\ -\frac{\underline{1}\cdot\underline{X}_{K-1}}{NT^2} & \cdots & 0 & 0 & \frac{1}{T} \end{bmatrix}$$

and $\alpha_\ell = -\frac{1}{N}[\frac{D_{1\ell}}{V_1} + \frac{D_{2\ell}}{V_2} + \cdots + \frac{D_{K-1,\ell}}{V_{K-1}}]$.

Since the matrix $[\underline{\underline{D}}]$ is now the inverse of a correlation matrix, by a well-known formula (Rao (1965), page 224)

$$(4.9) \qquad D_{\ell\ell} = \frac{\sigma_\ell^2}{\sigma_{\ell..}^2} .$$

From this we can evaluate the diagonal elements of $[\underline{\underline{X}}'\underline{\underline{X}}]^{-1}$ by reversing our operations with the elementary matrices $[\underline{\underline{F}}]$ and $[\underline{\underline{G}}]$ to establish that,

$$(4.10) \qquad [\underline{\underline{X}}'\underline{\underline{X}}]^{-1}_{\ell\ell} = \frac{D_{\ell\ell}}{\sigma_\ell^2 T} = \frac{1}{T\sigma_{\ell..}^2} \qquad \ell=1,2,\ldots,K-1$$

and

$$(4.11) \qquad [\underline{\underline{X}}'\underline{\underline{X}}]^{-1}_{00} = \frac{1}{N^2T} - \frac{1}{NT}(\frac{\alpha_1}{V_1} + \frac{\alpha_2}{V_2} + \cdots + \frac{\alpha_{K-1}}{V_{K-1}})$$

$$= \frac{1}{NT}[\frac{1}{N} - (\frac{\alpha_1}{V_1} + \frac{\alpha_2}{V_2} + \cdots + \frac{\alpha_{K-1}}{V_{K-1}})].$$

A typical term of (4.11) becomes

$$(4.12) \qquad \frac{1}{TN}\frac{\alpha_\ell}{V_\ell} = -\frac{1}{TN^2}(\frac{D_{1\ell}}{V_1V_\ell} + \frac{D_{2\ell}}{V_2V_\ell} + \cdots + \frac{D_{K-1,\ell}}{V_{K-1,}V_\ell}).$$

Now using the fact that $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ is a covariance matrix gives

$$(4.13) \qquad \frac{1}{N^2T}\frac{D_{j\ell}}{V_jV_\ell} = \frac{D_{j\ell}}{\sigma_j\sigma_\ell}\cdot\frac{\bar{X}_j\bar{X}_\ell}{TN^2} = \frac{1}{N^2}[\underline{\underline{X}}'\underline{\underline{X}}]^{-1}_{j\ell}\bar{X}_j\bar{X}_\ell \le \sqrt{[\underline{\underline{X}}'\underline{\underline{X}}]^{-1}_{jj}}\sqrt{[\underline{\underline{X}}'\underline{\underline{X}}]^{-1}_{\ell\ell}}\frac{\bar{X}_j}{N}\cdot\frac{\bar{X}_\ell}{N}$$

and condition (3.4a) implies that

$$(4.14) \qquad \frac{\bar{X}_\ell}{N} \le M \qquad\qquad \ell=1,\ldots,K-1$$

Therefore

$$(4.15) \qquad \left|\frac{1}{TN}\frac{\alpha_\ell}{V_\ell}\right| \le \max_{1\le j\le K-1} \frac{M^2K}{T} \frac{1}{\sigma_{\ell..}\sigma_{j..}}$$

and

$$(4.16) \qquad |(\underline{X}'\underline{X})^{-1}_{00}| \le \frac{1}{N^2T} + \max_{\substack{1\le j\le K-1\\ 1\le \ell\le K-1}} \frac{M^2K^2}{T} \frac{1}{\sigma_{\ell..}\sigma_{j..}} = \frac{1}{N^2T} + \frac{M^2K^2}{T} \frac{1}{\min\limits_{1\le\ell\le K-1} \sigma^2_{\ell..}}$$

which completes the proof of Theorem 2.

## 5. Limiting Properties of Macrovariances

The previous analysis has provided bounds for $\underline{\underline{V}}_N(\underline{b})$, the macro parameter variance in terms of $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$ and conditions on the microvariables and microparameter variances, when the macroparameters are defined simply as least squares estimates based on the macro data. We now discuss under what conditions $\underline{\underline{V}}_N(\underline{b})$ tends to zero as the number of elements in the aggregate increases.

It is clear from Theorem 1 that

$$\lim_{N\to\infty} \underline{\underline{V}}_N(\underline{b}) = 0 \tag{5.1}$$

if

$$\lim_{N\to\infty} N(\underline{\underline{X}}'\underline{\underline{X}})^{-1} = 0 \tag{5.2}$$

and conditions (3.4a) and (3.4b) are satisfied.

Theorem 3 implies that (5.2) will hold if

$$\lim_{N\to\infty} \frac{N}{T\sigma^2_{\ell..}} = 0, \qquad \ell=0,1,2,\ldots,K-1. \tag{5.3}$$

The multiple correlation cofficient, $R^2_{\ell..}$ is related to $\sigma^2_\ell$ by the identity $R^2_{\ell..} = 1 - \frac{\sigma_{\ell..}^{2}}{\sigma^2_\ell}$ and (5.3) can be rewritten as

$$\lim_{N\to\infty} \frac{N}{T\sigma^2_\ell(1-R^2_{\ell..})} = 0 \qquad \ell=0,1,2,\ldots,K-1 \tag{5.4}$$

In order for $(\underline{\underline{X}}'\underline{\underline{X}})$ to be invertible we must have $R^2_{\ell..} < 1$ for all $\ell$. In most applications related to economic data it is reasonable to assume that

a $\delta > 0$ exists so that

$$(5.5) \qquad \sup_N R^2_{\ell ..} < 1 - \delta$$

If (5.5) holds then having

$$(5.6) \qquad \lim_{N\to\infty} \frac{N}{T\sigma^2_\ell} = 0$$

is enough to imply (5.3).

Define the average variance among microvariables as

$$(5.7) \qquad s^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} (x_{i\ell t} - \bar{x}_{i\ell})^2$$

and the average simple correlation among microvariables as

$$(5.8) \qquad r = \frac{2}{N(N-1)} \sum_{i<j} \frac{1}{T} \sum_{t=1}^{T} \frac{(x_{i\ell t}-\bar{x}_{i\ell})(x_{j\ell t}-\bar{x}_{j\ell})}{s^2}$$

Then it follows that

$$(5.9) \qquad \frac{N}{T\sigma^2_\ell} = \frac{1}{Ts^2[1-(N-1)r]}$$

In any cases of conceivable interest, we would expect $s^2$ to be bounded away from zero. In many economic applications, r can be expected to be positive and bounded away from zero. (Since $\sigma^2_\ell > 0$ we must at least have $r \geq 1/(N-1)$.) A meaningful industrial aggregate is normally composed of firms with common

production methods and similar customers. While in the short run, one firm's gain may be another's loss, fluctuations in market demand will ordinarily be shared in proportion to each firm's productive capacity. While some firms grow in periods of declining demand and others fade when demand is growing, this "maverick" behavior is unlikely to dominate. In many cases it does appear that $(N-1)r$ will increase with $N$. Clearly, however, effectiveness from the point of view of aggregation gain depends on the strength of the average correlation among entities comprising the aggregate (as well as collinearity among the explanatory variables reflected in $R^2_{\ell\cdot\cdot}$). One might conjecture that the larger most aggregates become, even in a well designed aggregation procedure, the more dissimilar the components become, thereby putting definite limits on the amount of variance reduction that can in fact be achieved.

We have examined conditions which imply that $\underline{\underline{V}}_N(b) \to 0$ with increasing $N$. These conditions can be weakened if we only require that the elements of $\underline{\underline{V}}_N(b)$ remain bounded as $N$ becomes large. For example, because of the dissimilarities introduced as the number of components increases it might be that $(N-1)r$ in equation (5.9) is bounded (i.e. $r$ is not bounded away from zero). In this case $N/T\sigma^2_\ell$ would be bounded (if $s^2$ is bounded away from zero) and of course $\underline{\underline{V}}_N(b)$ would not necessarily approach zero. Since we do not observe almost zero coefficient variances even for very large $N$, the possibility raised here seems to be occurring in practice.

## 6. The Theil Weights

Theil (1954) originally showed that the macro parameters can be related to the microparameters through a set of aggregation weights which are constructs obtained through regressing each micro series on all of the macro exogenous variables, the resulting regression coefficients, denoted by the column vector $\underline{w}_i^{(\ell)}$, being the weights.* The formal relation is

$$(6.1) \qquad \underline{w}_i^{(\ell)} = (\underline{\underline{X}}'\underline{\underline{X}})^{-1}\underline{\underline{X}}'\underline{x}_i^{(\ell)},$$

when $\underline{x}_i^{(\ell)}$ $\ell=0,1,\ldots,K-1$ denote the columns of the T x KT matrix

$$\underline{\underline{x}}_i = \begin{bmatrix} \underline{x}_{i1} \\ \vdots \\ \underline{x}_{iT} \end{bmatrix} .$$

It follows immediately that

$$(6.2) \qquad \sum_{i=1}^{N} w_{ip}^{(\ell)} = \begin{cases} 1 & \text{for } p = \ell \\ 0 & \text{for } p \neq \ell \end{cases}$$

Thus, the weights for corresponding parameters sum to unity, and sum to zero for non-corresponding parameters. If we define a T x 1 vector of residuals

$$(6.3) \qquad \underline{u}_i^{(\ell)} = \underline{x}_i^{(\ell)} - \underline{\underline{X}}\underline{w}_i^{(\ell)}$$

then

---

*An accessible exposition of these concepts will be found in Klock (1961) and in a notation similar to that of this paper in Kuh (1972).

$$(6.4) \qquad \underline{x}_i^{(\ell)} = \underline{\underline{X}} \; \underline{w}_i^{(\ell)} + \underline{u}_i^{(\ell)}.$$

It is now possible to use these weights to relate the macro and micro equations. A few definitions are necessary:

$$\underline{\underline{U}}_i = [\, \underline{u}_i^{(0)} \;\; \underline{u}_i^{(1)} \; \ldots \; \underline{u}_i^{(K-1)} \,] \qquad (T \times K)$$

$$(U_i)_t = \text{t-th row of } U_i \qquad t=1,2,\ldots,T$$

$$\underline{\underline{Z}} = \begin{bmatrix} \sum_{i=1}^{N} \underline{x}_{i1} & \underline{0} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \sum_{i=1}^{N} \underline{x}_{i2} & \underline{0} & \cdots & \underline{0} \\ \vdots & \vdots & \sum_{i=1}^{N} \underline{x}_{i3} & \cdots & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots \sum_{i=1}^{N} & \underline{x}_{iT} \end{bmatrix} \qquad (T \times KT)$$

$$\underline{\underline{W}}_i = [\underline{w}_i^{(0)}, \underline{w}_i^{(1)}, \ldots, \underline{w}_i^{(K-1)}] \qquad (K \times K)$$

$$\underline{\underline{W}} = \begin{bmatrix} \underline{\underline{W}}_1 & \underline{\underline{0}} \cdots \underline{\underline{0}} & \underline{\underline{W}}_2 & \underline{\underline{0}} \cdots \underline{\underline{0}} & \cdots & \underline{\underline{W}}_N & \underline{\underline{0}} \cdots \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{W}}_1 \cdots \underline{\underline{0}} & \underline{\underline{0}} & \underline{\underline{W}}_2 \cdots \underline{\underline{0}} & \cdots & \underline{\underline{0}} & \underline{\underline{W}}_N \cdots \underline{\underline{0}} \\ \vdots & \vdots \quad \vdots & \vdots & \vdots \quad \vdots & & \vdots & \vdots \quad \vdots \\ \underline{\underline{0}} & \underline{\underline{0}} \quad \underline{\underline{W}}_1 & \underline{\underline{0}} & \underline{\underline{0}} \quad \underline{\underline{W}}_2 & & \underline{\underline{0}} & \underline{\underline{0}} \quad \underline{\underline{W}}_N \end{bmatrix} \qquad (KT \times KTN).$$

Then we have

$$(6.5)\qquad \underline{Y} = \underline{\underline{Z}}\ \underline{\underline{W}} \begin{bmatrix} \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_N \end{bmatrix} + \tilde{\underline{\varepsilon}}$$

and

$$(6.6)\qquad \tilde{\underline{\varepsilon}} = \begin{bmatrix} \sum_{i=1}^{N} [(\underline{\underline{U}}_i)_1\ \underline{\beta}_{i1} + \varepsilon_{i1}] \\ \vdots \\ \sum_{i=1}^{N} [(\underline{\underline{U}}_i)_T\ \underline{\beta}_{iT} + \varepsilon_{iT}] \end{bmatrix} \qquad T \times 1$$

We can also relate these weights to the matrix $(\underline{\underline{X}}'\underline{\underline{X}})^{-1}$.

Theorem 3. If (3.4 a) holds then

$$(6.7)\qquad \sum_{i=1}^{N} (w_{ip}^{(\ell)})^2 \leq M^2TN[(\underline{\underline{X}}'\underline{\underline{X}})^{-1}_{pp}] .$$

Proof. The proof is essentially the same as that used for Theorem 1.

$$(6.8)\qquad \sum_{i=1}^{N} (w_{ip}^{(\ell)})^2 = \sum_{i=1}^{N} \underline{g}^{(p)}\underline{x}_i^{(\ell)}\underline{x}_i'^{(\ell)}\underline{g}'^{(p)}$$

$$= \mathrm{tr} \sum_{i=1}^{N} \underline{x}_i^{(\ell)}\underline{x}_i'^{(\ell)}\underline{g}'^{(p)}\underline{g}^{(p)}$$

$$\leq NM^2 \mathrm{tr}\ 1|\underline{g}'^{(p)}||\underline{g}^{(p)}| .$$

The remaining steps are identical to (3.9) and (3.10) in the proof of Theorem 1.

We have seen in Sections 3 and 5 how aggregation gain is related to $N(\underline{X}'\underline{X})^{-1}$. If we require that (5.2) hold then Theorem 3 implies that

$$(6.9) \qquad \lim_{N\to\infty} \sum_{i=1}^{N} (w_{ip}^{(\ell)})^2 = 0$$

and as a consequence

$$(6.10) \qquad \lim_{N\to\infty} w_{ip}^{(\ell)} = 0 \qquad\qquad \ell=1,\ldots,K-1;\ p=1,\ldots,K-1$$

But (6.2) states that

$$(6.11) \qquad \sum_{i=1}^{N} w_{i\ell}^{(\ell)} = 1$$

for all n.

If we view the diagonal weights $w_{i\ell}^{(\ell)}$ as proportions, representing the relative contribution of the micro components $\underline{x}_{i\ell}$ to $\underline{X}_{\ell}$ then a necessary condition for aggregation gain is that no micro explanatory factor can be a large proportion of the aggregate of that factor.

The earlier paper by Kuh (1972), using data for four industries, found that corresponding aggregation weights frequently resemble proportions as one might expect. An approximate empirical measure of aggregation gain that ignores the non-corresponding weights (which typically are small) can be defined. It is the reciprocal of the sum of squares of the corresponding weights, (or alternatively of the sum of squares of proportional shares) which should indicate the reduction in the macro parameter variance relative to the average micro parameter variance. The empirical results showed that aggregation gain usually does occur and that the predictive value of

the measures of aggregation gain do reasonably well although not uniformly so. Further theoretical and empirical work on aggregation benefits and losses in terms of $N(\underline{X}'\underline{X})^{-1}$ will be pursued in subsequent work.

## References

[1] Burnett, T. and Guthrie, D. (1970). Estimation of Stationary Stochastic Regression Parameters. JASA. 65, pp. 1517-1573.

[2] Klock, T. (1961). Note on a Convenient Notation in Multivariate Statistical Analysis and in the Theory of Linear Aggregation. International Economic Review. 2, pp. 351-360.

[3] Kuh, E. (1972). An Essay on Aggregation Theory and Practice in Essays in Honor of Jan Tinbergen (W. Sellekaerts, editor), New York: Macmillan, (forthcoming).

[4] Rao, C. R. (1965). Linear Statistical Inference. New York: Wiley.

[5] Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. Econometrica. 38, pp. 3-23.

[6] Theil, H. (1954). Linear Aggregation of Economic Relations. Amsterdam: North Holland Publishing Company.

[7] Theil, H. (1968). Consistent Aggregation of Micromodels with Random Coefficients. Report 6816, Chicago, Center for Mathematical Studies in Business and Economy (April 1968). (A more accessible and somewhat more extensive treatment can be found in H. Theil, Principles of Econometrics, New York: John Wiley and Sons, 1971).

BASEMENT


## Date Due

| | | |
|---|---|---|
| ~~NOV~~ | MAY | |
| MAY | | |
| FEB 20 '79 | 1992 | |
| MAR 22 '79 | 18 | |
| '79 | 199 | |
| ~~JUN 79~~ AUG 13 | Y 18 2000 | |
| | SEP 2017 | |
| '85 | | |
| JUN 3 '85 | | |
| SEP | | |
| NO 27 '87 | | |

Lib-26-67

MIT LIBRARIES
3 9080 003 701 866

BASEMENT

MIT LIBRARIES
3 9080 003 701 833

MIT LIBRARIES
3 9080 003 670 814

-J5 143 w no.614- 72
Velsch, Roy E/A modification of the N
36673 D*BKS 00027764
3 9080 000 747 904

MIT LIBRARIES
3 9080 003 670 863

-J5 143 w no.616- 72
Velsch, Roy E/The variances of regres
36653 D*BKS 00027762
3 9080 000 747 854

MIT LIBRARIES
3 9080 003 701 825

MIT LIBRARIES
3 9080 003 701 874

MIT LIBRARIES
3 9080 003 701 783

MIT LIBRARIES
3 9080 003 701 924

HD28.M414 no.623- 72
Merton, Robert/Fallacy of the log-norm
636656 D*BKS 00027760
3 9080 000 747 789

HD28.M414 no.619-72
Plovnick, Mark/Expanding professional
636663 D*BKS 00027761
3 9080 000 747 813